(((Haskins Laboratories)))

Word Knowledge and the Simple View of Reading

David Braze, Leonard Katz, James S. Magnuson, W. Einar Mencl, Whitney Tabor,
Julie A. Van Dyke,Tao Gong, Clinton L. Johns & Donald P. Shankweiler
braze@haskinslabs.org

(((Haskins Laboratories)))

I. Purpose

Gough and Tunmer's Simple View of Reading suggests that reading comprehension (RC) is the product of decoding skill (D) and oral language comprehension (LC)[1]. This study investigates the contentious claim that the Simple View requires enrichment through addition of an independent vocabulary factor [2,3]. In the context of the Simple View, vocabulary is just one aspect of LC, and so of equal import to comprehension of print or speech, as with all other aspects of LC.

The question under examination here is whether the role of vocabulary should be modeled as distinct from general language comprehension ability, as assessed by measures of listening comprehension. The question is important, because of the influence that the Simple View continues to have on theoretical proposals regarding the cognitive processes involved in reading and reading skill differences. It is essential to test the assumptions of the Simple View precisely because of its influence.

II. Method

PARTICIPANTS

Participant sample was drawn from low-literacy young adults (N=286; age=16-25y), and includes a wider range of such abilities than would typically be found in university students [2]. All received multiple tests of vocabulary, nonword reading, word reading, listening and reading comprehension. Raw scores for these variables are summarized in Table 1.

TABLE 1: Summary of Measures

Measures	N	Mean	SD	Min	Max
gm.rcomp	224	29.69	9.47	8.00	47.00
wj3.rcomp	226	33.32	4.27	21.00	43.00
piat.rcomp	286	26.52	7.69	2.00	41.00
towre.w	286	118.49	18.50	53.33	190.74
towre.nw	286	57.17	20.65	1.33	110.68
wj3.watt	286	23.94	6.08	2.00	32.00
wj3.wid	286	63.05	7.25	36.00	76.00
piat.lcomp	286	28.59	7.44	9.00	41.00
wj3.oralcomp	286	23.85	4.34	9.00	33.00
ppvt	285	159.89	21.24	107.00	198.00
wasi.vocab	286	45.89	13.85	12.00	78.00
age	286	20.18	2.34	16.06	24.98
edu	286	11.89	1.71	8.00	17.00

ANALYTIC APPROACH

Stepwise regression and confirmatory factor analysis (CFA) are used to analyze the data. Analyses were done using the R statistical environment [4,5]. Prior to latent variable analyses and hierarchical regression, we examined score distributions for normality and outliers. Skewness was observed in most variables and so Box-Cox transformations were applied across the board [6]. Transformed variables were subsequently standardized.

A few participants were not given two of three measures of reading comprehension. The Hawkins test of multivariate normality and heteroscedasticity is used to address the question of whether the essential characteristics of participants who received those tests are different from those of participants who did not [7]. The method relies on assessing homogeneity of covariances for groups with different patterns of missingness. Our data include 5 such patterns: 217 complete cases, 54 cases missing both WJIII and Gates reading comprehension, 6 missing WJIII reading comprehension only, 8 missing the Gates reading comprehension only, and 1 is missing the PPVT only. The latter is excluded from this test as Hawkins requires at least 2 cases in each group. In short, the Hawkins test indicates there is not sufficient evidence to reject the null hypothesis that missingness in our data set can be characterized as 'missing completely at random' ($p = 0.329$).

Latent variable models were fit with the R package lavaan [5] using full information maximum likelihood as the estimation method.

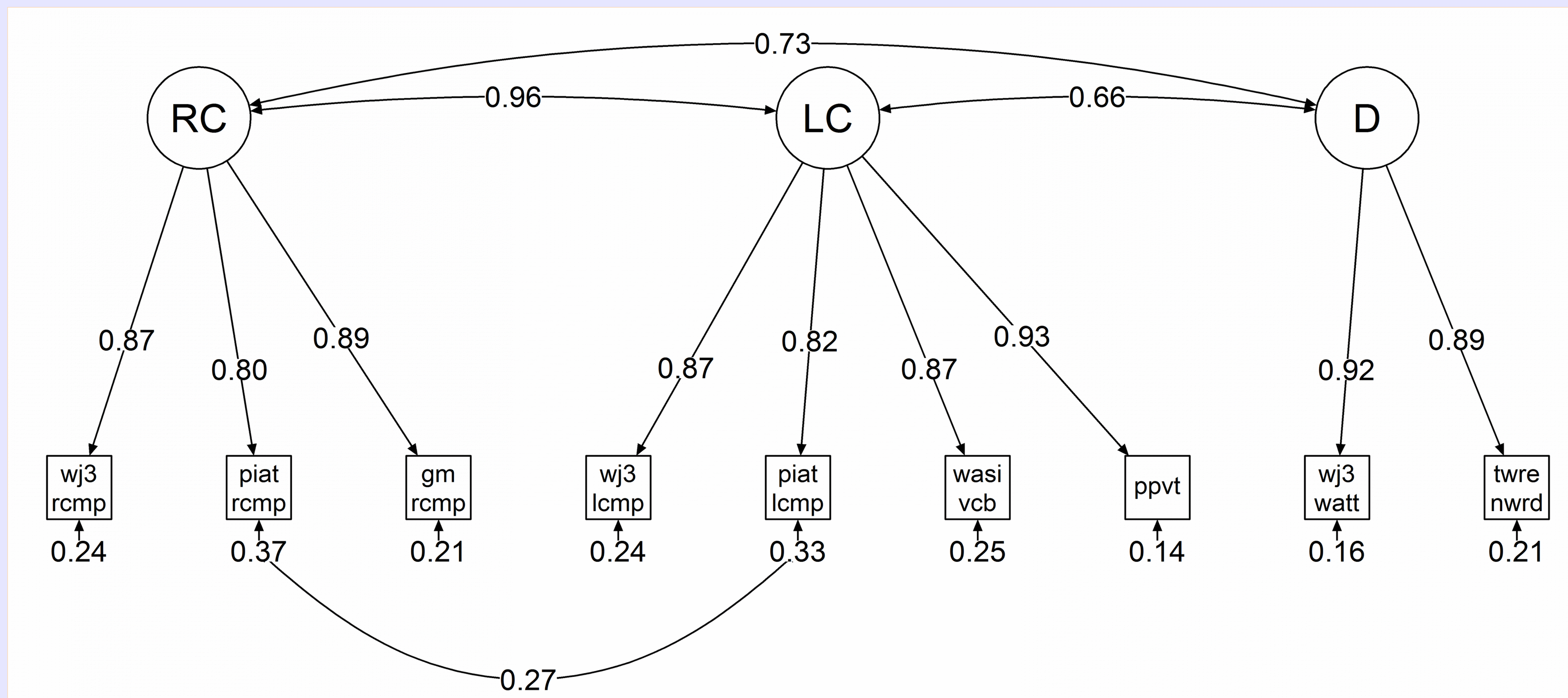
III. Result

TABLE 2: Correlations of transformed and standardized variables.

	1	2	3	4	5	6	7	8	9	10	11	12
1. gm.rcomp	--											
2. wj3.rcomp	.748	--										
3. piat.rcomp	.662	.689	--									
4. towre.w	.568	.545	.449	--								
5. towre.nw	.519	.543	.461	.765	--							
6. wj3.watt	.561	.610	.535	.605	.812	--						
7. wj3.wid	.703	.727	.657	.657	.761	.798	--					
8. piat.lcomp	.684	.664	.724	.442	.438	.480	.621	--				
9. wj3.oralcomp	.753	.720	.666	.437	.479	.503	.689	.717	--			
10. ppvt	.764	.764	.729	.500	.542	.589	.788	.774	.811	--		
11. wasi.vocab	.729	.701	.662	.483	.533	.557	.738	.698	.757	.805	--	
12. age	.075	-.014	.063	.070	.052	.069	.113	.113	.074	.110	.110	--
13. edu	.248	.275	.271	.268	.314	.279	.367	.331	.316	.403	.387	.464

There are strong correlations among measures of V and LC, with lesser correlations between these and measures of D.

CONFIRMATORY FACTOR ANALYSIS



The overall quality of fit suggests the model contains a reasonable set of latent constructs.

$\chi^2(24) = 21.21$
TLI = 1.0
RMSEA = 0.0
90% CI: 0 - .041

An alternative model with separate latent factors for V & LC failed to produce a positive definite matrix, a convergence problem likely due to the high correlations among the V and LC variables.

TABLE3: Model Implied Correlations (lower tri.) & Residuals (upper tri.)

	1	2	3	4	5	6	7	8	9	10	11
1. wj3.rcomp	--	.013	-.001	-.007	-.010	-.008	-.006	.033	-.007		
2. piat.rcomp	.697	--	-.008	-.005	-.001	-.004	.014	-.001	-.057		
3. gm.rcomp	.777	.708	--	.022	.007	.020	-.011	.007	-.013		
4. wj3.oralcomp	.737	.671	.749	--	.000	.001	-.002	-.030	-.036		
5. piat.lcomp	.692	.725	.704	.717	--	-.013	.010	-.020	-.046		
6. wasi.vocab	.730	.665	.742	.756	.711	--	-.001	.029	.023		
7. ppvt	.784	.714	.797	.812	.763	.805	--	.024	-.003		
8. wj3.watt	.588	.536	.598	.533	.500	.528	.567	--	-.002		
9. towre.nw	.568	.518	.577	.515	.483	.510	.547	.814	--		
10. RC	.874	.797	.889	.843	.792	.835	.897	.672	.650	--	
11. LC	.843	.768	.857	.874	.821	.866	.930	.610	.589	.965	--
12. D	.640	.584	.651	.580	.545	.575	.617	.918	.887	.733	.664

Uniformly low model residuals, the differences between empirical correlations and model implied correlations, indicate good local fit to the data, across the board.

TABLE4: Regression Models Targeting 3 Reading Comprehension Measures

	β	std.err	t.value	p.value	sspc
Gates-MacGinitie					
towre.nw	.048	.067	0.72	.474	.0007
wj3.watt	.144	.070	2.06	.041	.0061
piat.lcomp	.151	.063	2.39	.018	.0083
wj3.oralcomp	.296	.070	4.23	.000	.0259
ppvt	.180	.083	2.17	.031	.0068
wasi.vocab	.220	.078	2.84	.005	.0116
Woodcock-Johnson III	**β**	**std.err**	**t.value**	**p.value**	**sspc**
towre.nw	.008	.069	0.11	.912	.0000
wj3.watt	.213	.071	2.99	.003	.0138
piat.lcomp	.117	.066	1.77	.079	.0048
wj3.oralcomp	.220	.071	3.08	.002	.0147
ppvt	.280	.085	3.29	.001	.0168
wasi.vocab	.142	.079	1.80	.073	.0050
Peabody	**β**	**std.err**	**t.value**	**p.value**	**sspc**
towre.nw	-.069	.064	-1.07	.285	.0016
wj3.watt	.174	.066	2.62	.009	.0094
piat.lcomp	.358	.061	5.92	.000	.0482
wj3.oralcomp	.097	.067	1.44	.151	.0029
ppvt	.229	.080	2.86	.005	.0112
wasi.vocab	.092	.067	1.37	.172	.0026

Although CFA has already rejected the notion of V as distinct from LC, our intention here was to see if we could replicate, with present data, earlier results suggesting that some variance in RC can be attributed uniquely to V [2,3,8,9,10].

Specifically, we ask whether there is an increment in explained variance when measures of vocabulary are added to a regression model predicting reading comprehension from measures of decoding and oral language skill.

These models replicate earlier findings that vocabulary measures account for additional unique variance in reading comprehension after the contributions of word recognition and listening comprehension are accounted for.

IV. Conclusions

Latent variable modeling does not support the presence of a factor for V, independent of general oral language skill. Nonetheless, regression modeling indicates that measures of vocabulary do capture variance in reading comprehension beyond that captured by other oral language measures and decoding measures. As Tunmer and Chapman [2] point out, the variance in reading comprehension uniquely captured by vocabulary measures could arise as a result of low reliability in the predictor variables. The non-vocabulary variables used to represent LC may have failed to measure all aspects of LC relevant to RC and so measures of vocabulary may have captured additional aspects of LC not covered by the original variables. Moreover, different measures of RC may draw more or less heavily on various component reading skills, including word knowledge [11,12].

Beyond word recognition, improvement in reading comprehension will come from better integrating the word being read into the meaning of the phrase, sentence, or larger body of text. This integration is a function of the reader's ability to construct and update situation models, to retrieve information from both the text and general world knowledge, make critical inferences, establish coherence relations among text elements, and so on [13, 14], and is a major component of the reader's listening comprehension ability. All of this depends on the reader having high quality lexical representations, particularly in terms of their semantics, representations that can be flexibly integrated into higher level representations of discourse or narrative.

Thus, although our data indicate that, at least within the context of the Simple View of Reading, there is no basis for treating vocabulary as anything other than a component of listening comprehension, it is clear that vocabulary is an essential component of comprehension. Considerable evidence supports the idea that vocabulary knowledge is a limiting factor in reading comprehension [e.g., 15]. Moreover, we know how to teach it; it is a reading skill that has proved amenable to training [e.g., 16]. The importance of instruction in spoken vocabulary for preliteracy language development is well known and because preliteracy language skills mediate the efficacy of early reading instruction, high quality lexical representations remain a *sine qua non* of skilled reading.

References

1. Gough, P. B. & Tunmer, P. B. (1986). *Decoding, Reading, and Reading Disability*. Remedial and Special Education, 7, 6-10.
2. Braze, D., et al., (2007). *Speaking Up for Vocabulary: Reading Skill Differences in Young Adults.* Journal of Learning Disabilities, 40(3), 226-243.
3. Tunmer, W. E., & Chapman, J. W. (2012). *The Simple View of Reading Redux: Vocabulary Knowledge and the Independent Components Hypothesis*. Journal of learning disabilities, 45(5), 453-466.
4. R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
5. Rosseel, Y. (2012). *Lavaan: An R package for Structural Equation Modeling*. Journal of Statistical Software, 48(2), 1-36.
6. Box, G. E. P., & Cox, D. R. (1964). *An Analysis of Transformations*. Journal of the Royal Statistical Society. Series B (Methodological), 26(2), 211-252.
7. Jamshidian, M., & Jalal, S. (2010). *Tests of Homoscedasticity, Normality, and Missing Completely at Random for Incomplete Multivariate Data*. Psychometrika, 75(4), 649-674.
8. Fraser, J., & Conti-Ramsden, G. (2008). *Contribution of phonological and broader language skills to literacy*. International journal of language & communication disorders, 1-18.
9. Ouellette, G., & Beers, A. (2010). *A not-so-simple view of reading: how oral vocabulary and visual-word recognition complicate the story*. Reading and Writing, 23(2), 189-208
10. Protopapas, et al. (2013). *The Role of Vocabulary in the Context of the Simple View of Reading*. Reading & Writing Quarterly, 29(2), 168-202.
11. Cutting, L. E., & Scarborough, H. S. (2006). *Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills Can Depend on How Comprehension Is Measured*. Scientific Studies of Reading, 10(3), 277-299.
12. Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). *Reading Comprehension Tests Vary in the Skills They Assess: Differential Dependence on Decoding and Oral Comprehension*. Scientific Studies of Reading, 12(3), 281-300.
13. Hamilton, S. T., Freed, E. M., & Long, D. L. (2013). *Modeling reader and text interactions during narrative comprehension: A test of the lexical quality hypothesis*. Discourse Processes, 50(2), 139-163.
14. Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). *Low working memory capacity is only spuriously related to poor reading comprehension*. Cognition, 131(3), 373-403.
15. Perin, D. (2013). *Literacy Skills Among Academically Underprepared Students*. Community College Review, 41(2), 118-136.
16. Coyne, M. D., et al. (2010). *Direct & Extended Vocabulary Instruction in Kindergarten: Investigating Transfer Effects*. Journal of Research on Educational Effectiveness, 3, 93-120.

This research was supported by NIH grants R01 HD40353 & R01 HD071988 to Haskins Laboratories.